

Le nombre total de citations NTC d'un chercheur est informatif, mais son domaine de variation est assez large et il a plusieurs défauts qui le rendent inapte à opérer des comparaisons fines. Un chercheur ayant publié un seul travail dans un domaine central ou à la mode et qui, du coup, est très visible et largement cité sera considéré comme l'égal d'un autre qui aura publié une série de travaux difficiles et plus spécialisés, donc moins cités, alors que la contribution scientifique réelle du second peut être très supérieure. Un chercheur publiant des articles de synthèse sera favorisé, même s'il est peu

créatif. Les auteurs de livres seront en général défavorisés car, d'une part, les livres sont moins bien pris en compte que les articles dans les bases de données et, d'autre part, le travail pour écrire un livre est en général bien plus grand que pour un article.

L'indice h de J. Hirsch

C'est sans doute pour cela qu'en 2005, Jorge Hirsch, physicien de l'Université de Californie à San Diego, a introduit une nouvelle idée qui, depuis, est devenue le critère préféré pour évaluer les chercheurs. L'idée de l'indice

de Hirsch a donné lieu à tant d'études et de travaux de bibliométrie qu'aujourd'hui la publication de J. Hirsch de 2005 est citée 1 252 fois alors que sa publication la plus citée en physique ne l'est que 410 fois... ce qui est pourtant très bien !

L'indice h (ou h -index en anglais) de Hirsch prend en considération uniquement les meilleures publications d'un chercheur. L'indice h d'un chercheur est le plus grand entier k tel que les k articles les plus cités de ce chercheur ont chacun été cités au moins k fois. Pour avoir un indice h de 1, il faut avoir 1 publication citée 1 fois. Pour avoir un indice h de 2,

publications trop citées (par exemple en les fractionnant).

- Profil D. À nouveau un « chercheur linéaire », mais ses publications après l'intersection sont les plus inutiles et gâchent les citations (90 pertes après le point d'intersection, et 20 avant). Conseil : il doit se faire citer plus.
- Profil E. Ce chercheur est le plus « maladroit » de notre lot : 200 citations sont inutiles avant l'intersection,

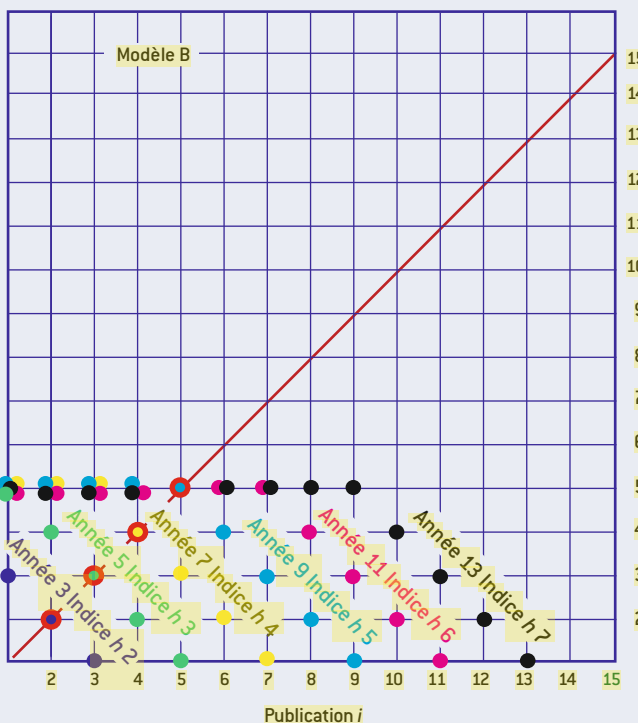
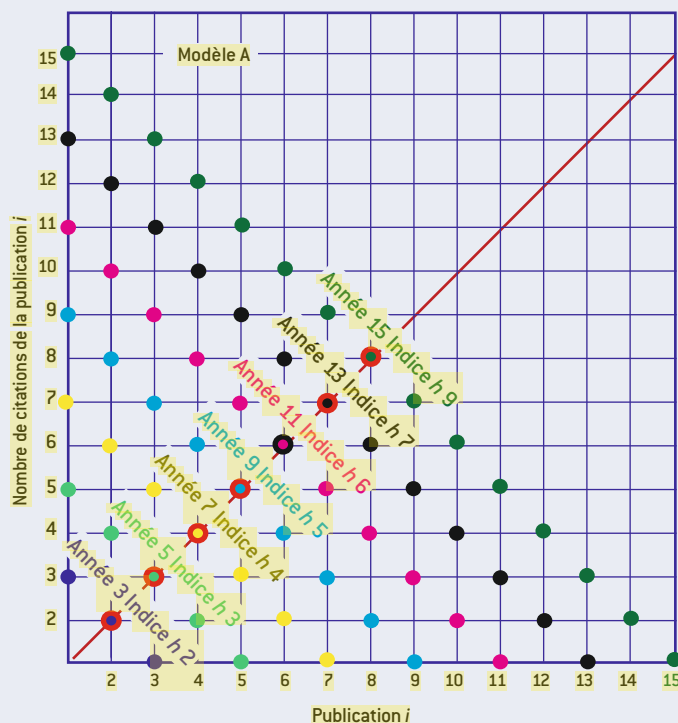
et 200 après. On imagine sans mal des cas pires encore.

Examinons maintenant l'évolution de l'indice h de chercheurs réguliers. Imaginons un chercheur (modèle A) qui publie un article chaque année et supposons que dès l'année où l'article a été publié, il est cité une fois par an exactement (y compris l'année de sa publication). Au bout d'un an, l'indice h

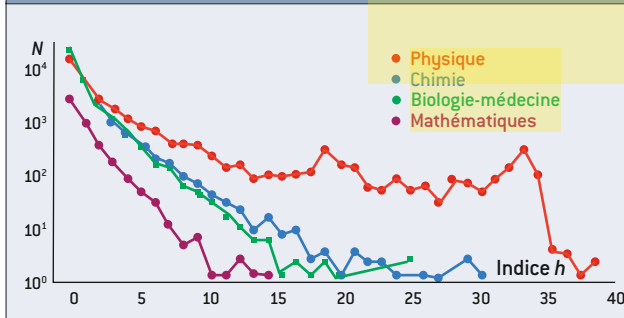
de notre chercheur imaginaire est 1 (il a un article cité une fois). Il lui faut attendre la fin de sa troisième année pour avoir un indice h de 2 et la fin de l'année $2k - 1$ pour avoir un indice h égal à k . L'indice h d'un tel chercheur croît linéairement.

L'hypothèse de citations régulières et indéfiniment prolongées pour chaque article publié est assez optimiste. En effet, les phénomènes de

mode et l'oubli ont pour conséquence qu'après avoir atteint un maximum, un article verra en général l'intérêt qu'il suscite diminuer. Si l'on suppose qu'au bout de cinq ans, on cesse de mentionner les articles de notre chercheur imaginaire (modèle B), cela aura un grave impact sur son indice h qui, après avoir atteint 5, cessera de croître, car aucun de ses articles ne sera jamais cité plus de cinq fois.



3. D'une discipline à l'autre



D'une discipline scientifique à l'autre, les usages en matière de publications et de citations sont très différents. Ainsi, les mathématiciens publient moins d'articles que les physiciens ou les biologistes et publient souvent seuls, ce qui n'est pas le cas en physique et encore moins en chimie. Le nombre de citations que reçoit en moyenne

un article varie aussi sensiblement d'une discipline à l'autre. L'indice h des prix Nobel de biologie vaut ainsi presque le double de celui constaté en physique. Le graphique ci-contre du nombre N de chercheurs brésiliens ayant un indice h donné est tiré d'un article de Pablo Batista *et al.* publié dans la revue *Scientometrics* en 2006.

il faut avoir au moins 2 publications citées au moins 2 fois chacune, etc.

L'indice h est donc majoré par le nombre de publications du chercheur : si vous n'avez publié que 5 articles, votre indice h ne pourra pas dépasser 5. Par ailleurs, si vous publiez beaucoup, mais que vous n'êtes jamais cité, votre indice h sera nul. Autre conséquence (voulu par J. Hirsch) de la définition : si vous avez un très bon NTC, mais que cela n'est dû qu'à un seul article, votre indice h vaudra 1.

Plus on avance, plus il est difficile de faire croître son indice h . Le nombre minimal de citations qu'il faut pour atteindre un indice h de k est k^2 . En général, on constate que le NTC d'un chercheur est 3 à 5 fois le carré de son indice h . Ce facteur est bien supérieur à 5 dans le cas où les meilleurs travaux sont beaucoup cités, ce qui est souvent le cas chez les chercheurs exceptionnels. Pour doubler son indice h , il faut être cité 4 fois plus, et pour le multiplier par 10, il faut que la notoriété des travaux soit multipliée par 100. Cet indicateur composite prend en compte l'intensité du travail d'un chercheur et sa qualité.

Voici un petit tableau pour les cinq mathématiciens mentionnés plus haut :

	N	NTC	Indice h
P. Erdős	975	15 040	60
N. Bourbaki	355	16 127	47
J. Dieudonné	310	12 580	39
A. Grothendieck	136	9 110	32
C. Villani	102	3 426	28

En dépit d'opposants déterminés, l'indice h constitue, selon les spécialistes de bibliométrie, une mesure efficace et assez robuste, qui donne une idée grossière et instantanée de la productivité et de la qualité du travail d'un chercheur. Cependant :

– Il ne faut pas oublier qu'il dépend de la base bibliographique permettant son calcul.

– Il peut y avoir des erreurs dues en particulier aux homonymies et aux difficultés typographiques. Il faut contrôler au moins les k premières publications pour un indice h valant k ; cet examen est plus facile que pour le NTC, car l'indice h ne dépend que d'un petit nombre de publications.

– Il est absurde de comparer des chercheurs de disciplines différentes, car les usages concernant les publications et les citations varient d'un domaine à l'autre. Les facteurs correctifs qu'on a tenté de calculer restent peu sûrs.

– De nombreux biais favorisent ou défavorisent un chercheur. En particulier, on doit prendre en compte l'âge d'un chercheur et le fait qu'il publie seul ou pas.

– Un seul nombre ne peut pas et ne pourra jamais évaluer à la fois la créativité, la persévérance, le rayonnement, l'influence, l'aptitude à la synthèse d'un individu. En conséquence, tout ramener à une seule note ne donnera jamais d'évaluation fine de la personnalité et des accomplissements précis d'un chercheur. C'est essentiel à retenir pour un comité qui doit recruter, attribuer une promotion ou décider d'un financement.

Dans son article de 2005, J. Hirsch a proposé une série d'arguments en faveur de son invention. Selon lui, parce qu'il limite l'importance des *hits* (publications citées de très nombreuses fois), l'indice h donne un classement plus juste des chercheurs que le NTC qui par ailleurs possède un pouvoir discriminant illusoire. De même, il n'a pas le défaut du nombre N de publications (car un chercheur qui publie beau-

coup sans jamais être cité n'est probablement pas un bon chercheur).

L'indice h est un meilleur indicateur que le nombre d'articles cités au moins K fois, qui décompterait aussi les bons articles d'un chercheur, mais dépendrait d'un facteur arbitraire K . Ce n'est pas le cas de l'indice h .

L'indice h d'un chercheur ne peut qu'augmenter. Avec des hypothèses simples sur la productivité d'un chercheur et l'écho de son travail, on aboutit à un modèle théorique de chercheur ayant un indice h qui croît linéairement tout au long de sa carrière, ou alors qui croît jusqu'à un maximum que le chercheur ne dépasse plus (voir la figure 2).

Du pour et du contre

D'après J. Hirsch, l'indice h est peu sensible (du moins quand il n'est pas trop petit) au phénomène de l'autocitation et aux petites erreurs que contient nécessairement une base de données bibliographique.

En examinant par exemple une série de lauréats du prix Nobel de physique, J. Hirsch a constaté qu'ils possèdent tous un indice h supérieur à 20. D'autres tests ont confirmé dans d'autres disciplines que les personnalités les plus en vue ont tous de bons indices h (pour leur discipline).

J. Hirsch a aussi mené une étude statistique pour évaluer la capacité prédictive de l'indice h . Prenant en compte l'indice h au bout de 12 années de recherche d'une série de scientifiques, il a calculé la capacité de son indicateur à prévoir la qualité du travail dans les 12 années suivantes. Son étude semble concluante. Assez étrangement même, la corrélation entre les indices h au bout des